Stat 406 Spring 2010: homework 7

1 Confidence intervals do not correspond to the intuitive notion of "confidence"

Suppose X_1 and X_2 are iid with $p(X_i = \theta + 1) = p(X_i = \theta - 1) = 0.5$ for some fixed $\theta \in \mathbb{R}$. (For concreteness, let's say $\theta = 10$, although the particular value does not matter.) Consider the following confidence interval (which happens to be a single point, rather than an interval):

$$C(\mathcal{D}) = \begin{cases} \text{the point } 0.5(x_1 + x_2) & \text{if } x_1 \neq x_2 \\ \text{the point } x_1 - 1 & \text{if } x_1 = x_2 \end{cases} \tag{1}$$

1. Prove that $C(\mathcal{D})$ is a 75% confidence interval. Hint: do a case analysis of what data could be generated by the model, compute what $C(\mathcal{D})$ is for each possible data set, and with what probability.
2. Now suppose you get a data set where $x_1 = x_2$, say (9,9) or (10,10). What is the posterior probability $p(\theta|\mathcal{D})$, assuming a uniform prior for θ? What is the confidence interval?
3. Now suppose you get a data set where $x_1 \neq x_2$, say (9,11) or (11,9). What is the posterior probability $p(\theta|\mathcal{D})$, assuming a uniform prior for θ? What is the confidence interval?
4. Explain why the Bayesian and frequentist procedures differ. Which do you prefer and why?

2 p-values depend on irrelevant factors not present in the data

Suppose we have a new drug which we administer to $n = 14$ people of whom $f = 4$ remain sick (failures) and $s = 10$ recover (successes). We want to know the probability θ the drug makes people get well (success).

1. Suppose you learn that the data was collected by choosing to measure the outcomes on $n = 14$ people. In this case, n is fixed and s (and hence $f = n - s$) is random. In particular, $s \sim \text{Bin}(n, \theta)$, with the following pmf

$$\text{Bin}(s|n, \theta) = \binom{n}{s} \theta^s (1-\theta)^{n-s} \tag{2}$$

   Compute the posterior $p(\theta|n, s)$ under this likelihood, given a uniform prior for θ. (Just write down its formula; you do not need to specify its normalization constant.)

2. Compute a one-sided p-value for the null hypothesis that $\theta = 0.5$ under the binomial model. Can you reject the null at signifance level $\alpha = 0.05$? Hint: the cdf for a binomial in Matlab is called `binocdf`.
3. Now suppose you learn that the data was collected by choosing to measure until $f = 4$ people were sick. In this case, f is fixed and n (and hence $s = n - f$) is random. The probability model becomes the **negative binomial distribution**, which is a distribution over the positive integers 0, 1, 2, ... There are two common definitions for this. In the first, we define the distribution over the number of trials n which will occur until we observe f failures. This has the form

$$\text{NegBinom}(n|f, \theta) = \binom{n-1}{f-1} \theta^{n-f} (1-\theta)^f \tag{3}$$

where θ is the probability of success. The reason for this formula is as follows: the last trial has to be a failure by definition, so we have $n-1$ places to choose between to allocate our $f-1$ other failures. An alternative definition is to define the negative binomial as the distribution over the number of successes s which will occur until we observe f failures. This has the form

$$\text{NegBinom}(s|f,\theta) = \binom{s+f-1}{f-1}\theta^s(1-\theta)^f \quad (4)$$

Since $n = s + f$, these are equivalent.

Compute the posterior $p(\theta|s,f)$ under this likelihood, given a uniform prior for θ.

4. Compute a one-sided p-value for the null hypothesis that $\theta = 0.5$ under the negative binomial model. Now can you reject the null at significance level $\alpha = 0.05$? Hint: the cdf for a negative binomial in Matlab is called `nbincdf`.

5. Explain why the Bayesian and frequentist procedures differ. Which do you prefer and why?

Turn in your numbers, calculations and any matlab code.

3 Bayesian analysis of the exponential distribution

A lifetime X of a machine is modeled by an exponential distribution with unknown parameter θ. The likelihood is $p(x|\theta) = \theta e^{-\theta x}$ for $x \geq 0$, $\theta > 0$.

1. Show that the MLE is $\hat{\theta} = 1/\overline{x}$, where $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

2. Suppose we observe $X_1 = 5, X_2 = 6, X_3 = 4$ (the lifetimes (in years) of 3 different iid machines). What is the MLE given this data?

3. Assume that an expert believes θ should have a prior distribution that is also exponential

$$p(\theta) = \text{Exp}(\lambda) \quad (5)$$

He believes the expected lifetime is $1/3$ of a year. What value of the prior parameter λ encodes this belief? (Call it $\hat{\lambda}$.) Hint: recall that the Gamma distribution has the form

$$\text{Ga}(\theta|a,b) \quad \propto \quad \theta^{a-1}e^{-\theta b} \quad (6)$$

and its mean is a/b.

4. What is the posterior, $p(\theta|\mathcal{D},\hat{\lambda})$?

5. Is the exponential prior conjugate to the exponential likelihood?

6. What is the posterior mean, $\mathbb{E}\left[\theta|\mathcal{D},\hat{\lambda}\right]$?

7. Explain why the MLE and posterior mean differ. Which is more reasonable in this example?